Manipulasi Frekuensi Dasar Menggunakan Metode STRAIGHT untuk Sintesis Suara Ucapan Ekspresif Dalam Bahasa Indonesia

Rizki Amalia F.K., Sekartedjo, dan Dhany Arifianto
Jurusan Teknik Fisika, Fakultas Teknologi Industri, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: sekar@ep.its.ac.id,dhany@ep.its.ac.id

Abstrak— ***Fundamental frequency* (F_0) merupakan salah satu parameter dalam sinyal suara ucapan yang dapat mempengaruhi tinggi rendahnya tekanan (intonasi). Parameter lain dalam sinyal suara ucapan yang berpengaruh terhadap intonasi adalah power, komponen periodik dan tak periodik sinyal suara. Dalam penelitian ini, dilakukan manipulasi sinyal suara ucapan hanya pada parameter F_0 sinyal suara ucapan berbahasa Indonesia, sedangkan parameter lain dianggap tetap. Manipulasi dilakukan dengan metode STRAIGHT. Kualitas hasil manipulasi suara ucapan dilakukan dengan metode MOS.**

Kata Kunci—***Fundamental Frequency* (F_0), *speech morphing*, STRAIGHT, sintesis suara ucapan ekspresif bahasa Indonesia.**

I. PENDAHULUAN

PERKEMBANGAN perangkat teknologi yang saat ini banyak dirasakan manfaatnya dalam kehidupan sehari-hari, memungkinkan manusia sebagai pengguna teknologi melakukan interaksi atau komunikasi dengan perangkat teknologi itu sendiri. Hal ini berarti bahwa, perangkat teknologi semakin dituntut untuk memiliki kemampuan respon yang baik terhadap komunikasi yang dilakukan oleh pengguna. Salah satu respon yang dapat ditemukan saat ini adalah sintesis suara ucapan (*speech synthesis*) dari perangkat teknologi yang menyerupai suara manusia. Salah satu cabang dari *speech synthesis* adalah *speech morphing*.

Salah satu metode untuk manipulasi, analisis, serta sintesis suara ucapan adalah metode STRAIGHT, yang telah dikembangkan sejak tahun 1997. Sedangkan aplikasi metode STRAIGHT untuk memanipulasi suara ucapan sehingga terbentuk suara ucapan berintonasi atau dikenal dengan istilah *speech morphing* diperkenalkan pada tahun 2003 [1]. Pada penelitian sebelumnya, telah dilakukan manipulasi sinyal suara dalam pengucapan berbahasa Inggris dan Jepang. Manipulasi yang dilakukan adalah mengubah intonasi suara ucapan normal (datar) ke bentuk intonasi ekspresif. Manipulasi dilakukan pada parameter-parameter fisik sinyal suara ucapan, yaitu frekuensi dasar (F_0), *power*, komponen periodik serta komponan aperiodik [2,3].

Pada tugas akhir ini telah dilakukan penelitian tentang manipulasi sinyal suara dalam pengucapan berbahasa Indonesia menggunakan metode STRAIGHT. Penelitian dititik beratkan pada manipulasi F_0, sedangkan parameter fisik sinyal suara ucapan yang lain dianggap tetap. Tujuan dari tugas akhir ini adalah mendapatkan teknik penerapan metode STRAIGHT untuk memanipulasi suara ucapan dalam bahasa Indonesia, sehingga diperoleh suara ucapan ekspresif berbahasa Indonesia, serta mendapatkan kualitas hasil manipulasi sinyal suara ucapan secara subjektif dengan metode *Mean Opinion Score* (MOS) [4].

Tabel 1.
Daftar simbol fonetik dalam bahasa Indonesia

No.	Simbol	No.	Simbol	No.	Simbol
1	A	13	Ny	25	n
2	I	14	Sy	26	p
3	U	15	B	27	r
4	E	16	C	28	s
5	ê	17	D	29	t
6	O	18	F	30	w
7	Ai	19	G	31	y
8	Au	20	H	32	z
9	Ei	21	J		
10	Oi	22	K		
11	Kh	23	L		
12	Ng	24	M		

II. URAIAN PENELITIAN

A. Pembuatan Database Kalimat Bahasa Indonesia

Sebagai tahap awal untuk memulai penelitian sintesis suara ucapan berbahasa Indonesia, diperlukan *database* kalimat bahasa Indonesia yang harus memenuhi aturan kesetimbangan fonetik (*phonetically balanced*). Kesetimbangan fonetik yang dimaksud adalah terpenuhinya 32 simbol fonetik pada Tabel 1 dalam *database* kalimat bahasa Indonesia yang akan disusun [5].

B. Perekaman Database Kalimat Bahasa Indonesia dengan Intonasi Pengucapan Normal

Proses perekaman dilakukan di ruang kedap Laboratorium Akustik Jurusan Teknik Fisika. Yang menjadi subjek perekaman adalah narawicara, yaitu satu orang laki-laki dan satu orang wanita. Masing-masing narawicara melakukan perekaman beberapa kalimat bahasa Indonesia. Perangkat yang dibutuhkan untuk melakukan perekaman adalah :

1. Laptop yang telah terinstal software Adobe Audition 3.0

dan driver EMU.

Gambar 1. Sinyal suara ucapan hasil perekaman

2. *Microphone* dan *stand mic*
3. Headphone
4. *Audio coder* EMU.

File audio yang dihasilkan dari proses perekaman disimpan dalam bentuk rekaman suara ucapan berbahasa Indonesia, dalam format ".wav".

C. Perekaman Suara Ucapan Ekspresif Berbahasa Indonesia

Teknik perekaman suara ucapan ekspresif sama dengan teknik perekaman suara ucapan berintonasi normal. Narawicara harus dilatih untuk mengucapkan suara dalam beberapa ekspresi

D. Pengolahan Sinyal Suara

Hasil perekaman suara ucapan yang berupa file ".wav" dirapikan sebelum memasuki proses berikutnya. Perapian yang dilakukan yaitu memotong bagian akhir dan awal sinyal. Bagian akhir dan awal sinyal adalah bagian *silent*. Bagian *silent* merupakan bagian transisi, saat narawicara akan memulai proses perekaman dan saat narawicara mengakhiri proses perekaman.

Dapat dilihat pada Gambar 1, kotak berwarna merah menunjukkan bagian *silent* pada sinyal suara ucapan yang akan dipotong.

E. Ekstraksi F_0

Ekstraksi F_0 dilakukan dengan memanfaatkan menu ekstraksi F_0 pada STRAIGHT GUI Matlab. Hasil ekstraksi adalah nilai dan grafik F_0 sinyal suara.

F. Manipulasi F_0 Sinyal Suara Menggunakan STRAIGHT GUI

Manipulasi F_0 dilakukan untuk mengubah intonasi suara ucapan dari intonasi normal menjadi beberapa intonasi ekspresif. Manipulasi yang dimaksudkan adalah menggeser nilai F_0 awal sinyal suara ucapan berintonasi normal ke nilai tertentu yang lain secara manual. Sehinnga terjadi perubahan nilai pada titik dan daerah yang digeser.

G. Pengujian Subjektif dengan Mean Opinion Score (MOS)

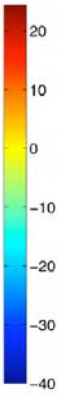
MOS merupakan suatu metode yang digunakan untuk mengukur kualitas suara. Sedangkan objek uji subjektif disebut sebagai naracoba. Adapun standar naracoba yang telah ditentukan dalam MOS adalah:

1. Pendengar belum terlibat secara langsung dalam pekerjaan yang akan diujikan
2. Pendengar tidak berpartisipasi pada subjektif tes apa pun dalam kurun waktu enam bulan sebelumnya dan tes opini-pendengaran apa pun setidaknya satu tahun, serta
3. Pendengar tidak pernah mendengar daftar kalimat yang sama sebelumnya.

Masing-masing naracoba melakukan penilaian terhadap intonasi sinyal suara ucapan hasil manipulasi yang

Tabel 2.
Parameter Penilaian uji MOS

Parameter Kualitas	*Score*
Sangat baik	5
Baik	4
Cukup baik	3
Buruk	2
Sangat buruk	1



Gambar 3. Tingkat warna pada spektogram.

direpresentasikan dalam bentuk *score*, dapat dilihat pada Tabel 2 [4]. *Score* subjektif seluruh naracoba tersebut akan dihitung rata-ratanya sehingga mendapatkan *score* opini rata-rata.

H. Hasil Manipulasi F_0 Sinyal Suara.

Hasil manipulasi F_0 pada suara ucapan berbahasa Indonesia dari intonasi normal ke intonasi ekspresif dapat ditampilkan dalam bentuk spektogram. Dalam spektogram terdapat warna-warna yang merepresentasikan tingkat *power* suara pada frekuensi dan waktu tertentu. Warna merah menunjukkan tingkat power tertinggi (10-20 dB). Dimana warna merah tersebut merupakan power suara ucapan. Perbedaan tingkat *power* tertinggi hingga terendah ditunjukkan dengan warna merah dan biru. Tingkat power tertinggi dengan nilai 20dB, sedangkan tingkat power terendah dengann nilai -40dB.

Tidak terjadinya perubahan yang diakibatkan proses manipulasi F_0 membuktikan bahwa dengan metode STRAIGHT, manipulasi yang dilakukan pada salah satu parameter sinyal suara dapat dilakukan secara independen tidak mempengaruhi inkonsistensi pada parameter-parameter fisik yang lain [5].

I. Hasil Uji MOS

Penilaian secara subjektif oleh beberapa orang naracoba terhadap hasil manipulasi, menunjukkan bahwa kualitas suara ucapan ekspresif hasil manipulasi memiliki kualitas yang cukup baik.

III. KESIMPULAN

Kesimpulan yang dapat diperoleh dari penelitian yang telah dilakukan adalah

1. Telah didapatkan teknik penerapan metode STRAIGHT

untuk memanipulasi sinyal suara ucapan berbahasa Indonesia.

2. Kualitas hasil manipulasi suara ucapan ekspresif yang diperoleh dari uji MOS yaitu cukup baik.

DAFTAR PUSTAKA

[1] H. Kawahara, H. Matsui, "Auditory Morphing Based on an Elastic Perceptual Distance Metric in an Interference-Free Time-Frequency Representation", ICASSP'2003, Hongkong, 2003, vol. 1, pp. 256-259.

[2] H. Kawahara, "STRAIGHT, Exploitation of the other of VOCODER: Perceptually Isomorphic Decomposition of Speech Sound", Acoust, Sci & Tech. 27, 6, 2006.

[3] http://www.wakayama-u.ac.jp/~kawahara/Miraikandemo/straightMorph.swf

(Demo apllikasi STRAIGHT)

[4] http://www.itu.int/rec/T-REC-P.800-199608-I/en

(Manual informasi uji MOS)

[5] Suyanto, "An Indonesian Phonetically Balanced Sentence Set for Collecting Speech Database", Jurnal Teknologi Industri Vo. XI No. 1Januari 2007: 59-68, 2007.